WHAT IS CLAIMED IS:

1. In a network, a method for segmenting a streaming multimedia clip into a plurality of sequentially organized data segments of exponentially increasing size and distributing said streaming multimedia clip from an origin server to a plurality of streaming caches which comprise a distribution set in said network, the method comprising the steps of:

determining a size (L) of the multimedia clip;

segmenting the streaming multimedia clip into a plurality of data segments of exponentially increasing size; and

distributing the plurality of data segments from the origin server to said plurality of streaming caches, wherein an i-th data segment is distributed in an i-th distribution round to each of said plurality of streaming caches.

2. The method according to Claim 1, wherein the size of an i-th data segment is computed as $L / 2^{(N+1-i)}$

where N is the total number of segments, and

where i is an index defining each of the N segments, (i= 1,2,….,N).

3. The method according to Claim 1, wherein the size L of the multimedia clip is measured in units of time.

4. The method according to Claim 1, wherein the segmenting step further comprises the steps of:

determining in an m-th distribution round if a data segment of said multimedia clip is equal to or greater than a predetermined threshold value, said m-th data segment referred to as a threshold data segment; and

dividing a remaining undivided portion of said multimedia clip into data segments having a predetermined segment size if the data segment of said multimedia clip is equal to or greater than a predetermined threshold value.

5. The method according to Claim 4, wherein said remaining undivided portion is divided into data segments in successive rounds having an index m+1 through N.

6. The method according to claim 4 wherein the predetermined segment size is equal to the size of the threshold data segment.

7. The method according to claim 4 wherein the predetermined segment size is computed as:

$$2^{(r-1)} * \delta$$

where $\delta = L / 2^{(N-1)}$ the size of a first segment; and

where r is a user adjustable parameter to determine the segment size for those fixed

segments which occur once the predetermined threshold has been reached.

8. The method according to Claim 5, wherein δ is on the order of 5 to 30 seconds.

9. The method according to Claim 4, wherein the values for δ, r and m are determined by an origin server in accordance with an origin server aware scheme.

10. The method according to claim 4, wherein the values for δ, r and m are determined by inter-cache communications in an origin server transparent scheme.

11. The method of Claim 1, wherein the distributing step further comprises the step of:

at each of said plurality of streaming caches, storing an i-th data segment of said streaming multimedia clip with probability equal to $1/2^{(i-1)}$ in an i-th distribution round, where $i = 1, 2, \ldots, N$.

12. A method for segmenting a streaming multimedia clip into a plurality of sequentially organized data segments of exponentially increasing size, comprising the steps of:

determining a size (L) of the streaming multimedia clip; and

segmenting the streaming multimedia clip into a plurality of data segments of exponentially increasing size.

13. A method of distributing a segmented streaming multimedia clip among a plurality of streaming caches, comprising the steps of:

at each of said streaming caches:

storing an i-th data segment of the segmented streaming multimedia clip with probability equal to $1/2^{(i-1)}$.

14. The method according to Claim 13, further comprising the step of:

storing an i-th data segment of said segmented streaming multimedia clip with probability equal to $[1/2^{(i-1)}] * e(x)$, where $e(x)$ is a constant that is proportional to a popularity rating of the clip, where $0 \leq e(x) \leq 1$.

15. A method of replacing segments in an SC, the method comprising:

(a) computing a potential function for each stored segment in said SC;

(b) sorting said stored segments into one of a plurality of bands wherein said bands are organized from a highest order band to a lowest order band, said bands being defined by an upper and a lower band boundary having values corresponding to the potential function; and

(c) replacing segments as needed in a sequence starting from those segments stored in said lowest order band to said highest ordered band, wherein segments are replaced in each band starting with segments having a corresponding lowest potential function value.

16. The method of Claim 14, wherein said potential function is computed for a segment j of clip i as:

F(i,j) = Prob (selecting a clip with rank i) * Prob (selecting segment j of the clip)

Where rank is determined using a global clip hotness rating.

17. The method of Claim 15, wherein said potential function is quantized prior to said sorting step.

18. A method of replacing segments in an SC, the method comprising:

(a) identifying a multimedia clip in said SC having a lowest global clip hotness rating;

(b) in the case where it is determined that said identified multimedia clip's global hotness rating is lower than a first threshold;

(1) removing said identified multimedia clip from said SC; and

(2) repeating steps (a)-(b) until either a sufficient amount of disk space is freed to terminate said method or step (b) is not satisfied;

(c) if step (b) is not satisfied, removing a number of segments of said identified multimedia clip from said SC starting from a last segment until either a sufficient amount of disk space is freed to terminate said method or a predefined threshold percentage of

said identified multimedia clip remains; and

(d) if the predefined threshold percentage of the clip remains, identifying a multimedia clip in said SC having the next lowest global clip hotness rating, and repeating step (c).

19. A system for segmenting, distributing and replacing segments of streaming multimedia clips in a network, comprising:

at least one origin server storing said streaming multimedia clips;

a plurality of streaming caches in communication with said at least one origin server, said plurality of streaming caches defining a distribution set;

first processing means associated with said at least one origin server for segmenting the streaming multimedia clip into a plurality of data segments of exponentially increasing size and for distributing said plurality of data segments to each of said plurality of streaming caches; and

second processing means associated with each of said plurality of streaming caches for storing data segments received from said at least one origin server in a SC and for replacing said stored data segments from said SC.

20. The system of Claim 19 wherein said second processing means further comprises means for computing a potential function for each stored data segment for replacing segments.

21. The system of Claim 19 wherein said second processing means further comprises means for computing a probability to determine whether to store or discard each data segment received from said at least one origin server.